

The University of Chicago

STATISTICAL ANALYSIS
OF PERSONALITY RATINGS

A PART OF A DISSERTATION SUBMITTED
TO THE FACULTY OF THE DIVISION OF
THE SOCIAL SCIENCES IN CANDIDACY FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATION

1936

By
PAN-LIN CHI

Private Edition, Distributed by
THE UNIVERSITY OF CHICAGO LIBRARIES
CHICAGO, ILLINOIS

Reprinted from
THE JOURNAL OF EXPERIMENTAL EDUCATION
Vol. V, No. 3, March 1937

The University of Chicago

STATISTICAL ANALYSIS
OF PERSONALITY RATINGS

A PART OF A DISSERTATION SUBMITTED
TO THE FACULTY OF THE DIVISION OF
THE SOCIAL SCIENCES IN CANDIDACY FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATION

1936

By
PAN-LIN CHI

Private Edition, Distributed by
THE UNIVERSITY OF CHICAGO LIBRARIES
CHICAGO, ILLINOIS

Reprinted from
THE JOURNAL OF EXPERIMENTAL EDUCATION
Vol. V, No. 3, March 1937



STATISTICAL ANALYSIS OF PERSONALITY RATING

PAN-LIN CHI
34 Hsing Hua Sih Street
Peiping, China

PART I CONSISTENCY

INTRODUCTION

Purpose of Study. Rating is one of the most widely used methods of studying character and personality. Since its procedure has not been well standardized, it has been used under various conditions. Raters have varied from casual novices to trained experts. Traits employed have varied from those which may be readily observed to those which are nearly abstract. Subjects rated have varied from infants to adults. And scales have varied from a graphic description of several definite categories to ranking in terms of relative positions in a group. Results obtained under one set of conditions are usually different from those obtained under another set of conditions. Thus the value of this method is debatable.

The purpose of this study is to ascertain the consistency and significance of the personality ratings made by teachers of their pupils in the Elementary School of the University of Chicago. The conditions under which the ratings were made were in some respects more favorable than those under which the ratings of previous studies were made: (1) the teachers were generally better trained in rating methods; (2) they had closer observation and better acquaintance with the pupils; (3) the pupils were in the fifth and sixth grades, and they probably express their feelings and attitudes with less restraint than do adults; and (4) the traits for rating were defined in terms of actions which are perhaps more accessible for observation than are the abstract names of traits. The analysis of these ratings may be expected to throw some light upon the real value of rating methods in general.

¹This article is a part of a dissertation submitted to the Department of Education of the University of Chicago in candidacy for the degree of Doctor of Philosophy.

The writer is grateful to Professor F. N. Freeman and to Professor K. J. Holzinger for the suggestions and criticisms they gave throughout this study.

This paper consists of three parts: the first is concerned with the consistency of the ratings; the second is a study of the halo effect; and the last part is a factor analysis.

Sources of Data. This study is based upon the ratings of 100 pupils in two classes in the Elementary School. The pupils of each class were rated on personality traits by their teachers in the fifth grade and also in the sixth grade. These records are available in the Records Office of the Department of Education of the University.

The first class of pupils entered the fifth grade in the fall of 1932 and the sixth grade in the autumn of 1933. The second class was in the fifth grade during the school year 1933-1934 and in the sixth grade during the following year. Upon the elimination of those pupils who did not stay long enough to have all the rating records for the two years, complete data for one hundred pupils remained.

Although the one hundred pupils did not pass each of the two grades at the same time, their data are comparable, as they were rated by the same persons in each of the two grades.

Each of the pupils was rated, during the entire period, by five teachers, two of whom taught in the fifth grade, while the other three taught in the sixth grade. One of the fifth grade teachers taught social studies, and the other English and arithmetic. Of the sixth grade teachers, one taught French, the second social studies, and the third English and arithmetic. One teacher made two ratings a year, or one rating each semester. There were ten ratings for each pupil in the two years.

The personality scale used consisted of twenty-six traits, each of which was defined in terms of several actions (ranging from one to eight). The pupils were rated with respect to the actions. Each action was analyzed on a graphic scale of five categories, namely, "never," "rarely," "occasionally," "frequently," and "consistently." In the present

investigation, the five categories have been designated by the five numerals "5," "4," "3," "2," and "1," with "5" for "never" and "1" for "consistently." Thus the ratings have been transformed from qualitative into quantitative statements for the convenience of statistical treatment.

Only nineteen of the twenty-six traits have been used in this study. They are attentiveness, mental alertness, intellectual curiosity, originality, industry, persistence, accuracy, carefulness, neatness, promptness, initiative, independence, leadership, self-assertion, facing reality, co-operation, sociability, emotional balance, and muscular coordination. The other traits were discarded, because the ratings concerning them either were not discriminative or were incomplete.

The rating on each of the nineteen traits was obtained by averaging the estimates regarding the actions in terms of which it was defined. The composite rating of one pupil was obtained by averaging his ratings on the nineteen traits. The rating of one pupil for one year is the average of his two ratings for the two semesters of the year. In this way the data for each pupil were recorded on the nineteen traits and for their composite, for each semester and also for each of the two years.

METHODS OF INVESTIGATION

The consistency of the ratings of the present study has been determined in two ways: first, in terms of the correlation between the rating and re-rating on each of the nineteen traits and their composite by each of the five teachers; secondly, in terms of the correlations between every two of the five teachers. The former indicates how consistently each teacher judges; while the latter determines to what extent one teacher agrees with another in judging the same traits.

Lastly, the correlation between the rating by each of the five teachers and the consensus of the other four has been computed. This correlation is sometimes used as a validity coefficient for evaluating ratings.

The correlation coefficients throughout this study have been calculated by the product-moment method, and for the convenience of computation Holzinger's formula 33 has been used.²

RELIABILITY BETWEEN RATING
AND RE-RATING

The reliability coefficient between rating and re-rating by each of the five teachers with respect to each of the nineteen traits and their composite is presented in Table I.

² Karl J. Holzinger, *Statistical Methods for Students in Education*, p. 146. New York: Ginn and Co., 1928.

TABLE I
RELIABILITY COEFFICIENT OF EACH OF FIVE TEACHERS WITH RESPECT TO EACH OF NINETEEN TRAITS AND THEIR COMPOSITE

Trait	Teacher				
	A	B	C	D	E
Attentiveness.....	.750	.695	.892	.893	.850
Mental alertness.....	.858	.653	.904	.877	.795
Intellectual curiosity.....	.851	.789	.923	---	.678
Originality.....	.775	.816	.926	.836	---
Industry.....	.647	.609	.813	.867	.752
Persistence.....	.769	.617	.843	.868	.825
Accuracy.....	.800	.822	.694	.890	.831
Carefulness.....	.469	.739	.860	.846	.884
Neatness.....	.795	.834	.928	.855	.839
Promptness.....	.594	.711	.914	.732	.861
Initiative.....	.848	.673	.883	.902	.749
Independence.....	.421	.741	.858	.869	.679
Leadership.....	.805	.844	.955	---	.835
Self-assertion.....	.880	.821	.853	.911	.829
Facing reality.....	.817	.585	.879	.879	.778
Co-operation.....	.708	.766	.872	---	.854
Sociability.....	.937	.740	.933	.816	.859
Emotional balance.....	.836	.665	.933	.922	.823
Muscular coordination.....	.879	.768	.876	.911	.846
Composite.....	.924	.809	.924	.938	.888
Average.....	.770	.735	.883	.871	.813

*The data are not complete.

Among the five teachers, C shows the highest reliability, .883, while B shows the lowest reliability, .735, on the average. Nevertheless, the range of variation among the five averages, namely, .148, is rather small. The five teachers may therefore be regarded as equally reliable.

The average of the average reliability coefficients of the five teachers is .812. It is much higher than .650, the average of the reliability coefficients reported in three previous studies of similar nature.³ This seems to indicate that the favorable conditions under which the ratings were made have increased the reliability.

The coefficients of each teacher varied greatly with the traits; the highest was over twice as large as the lowest. In view of the fact that there was no common tendency for the teachers to show a high reliability on certain traits and a low reliability on others, this variation does not appear to be due solely to a difference in accessibility to observation of the various traits, which was suggested as the important factor in some of the previous studies. Another factor suggests itself as one cause of this variation, namely, a difference in attitude or interest of a teacher toward the various traits. That is, a teacher would show a high reliability on that trait in which she had a keen interest and consequently made a real effort in its observation and estimation. On the other hand, she would show a low reliability on another trait toward which she was indifferent and therefore tended to make no more than a perfunctory rating.

Furthermore, the teachers differed markedly in the reliability of judgment as regards one trait. For instance, as shown in Table I, teacher A showed a reliability of .421 for independence, while D showed .869 for the same trait; for emotional balance, B showed .665, while C showed .933; for intellectual curiosity, E showed .678, while C showed .923; and so on. This variation also may be accounted for by a difference in attitude of the teachers toward a trait. Theoretically, when one teacher obtains a high reliability on one trait, it should be possible for another teacher to obtain the same reliability.

³ W. Hardin Hughes, "Relation of Intelligence to Trait Characteristics," *Journal of Educational Psychology*, XVII (1926), 482-94.

A. W. Kornhauser, "A Comparison of Raters," *Journal of Personnel Research*, V (1927), 338-44.

H. S. Conrad, "The Validity of Personality Ratings of Pre-school Children," *Journal of Educational Psychology*, XXII (1932), 671-80.

Another fact which may support the foregoing conclusion is that the traits vary only slightly in the reliability of rating, as measured by the average of the five reliability coefficients of the five teachers. As shown in Table II, the average coefficients of reliability of the traits range from .714 for independence to .877 for the composite, with an average of .812. The range of variation, .146, is rather small. This seems to indicate that the traits, after being specifically defined, can be estimated with equal reliability, so far as rating and re-rating by one teacher is concerned.

When the average of the five ratings for the first semester is correlated with the average of the five ratings for the second semester on each of the nineteen traits and the composite, the coefficients, as shown in Table III, range from .856 for originality to .946 for accuracy, with an average of .909. This average is higher than the average reliability of each of the five teachers, which ranged, as mentioned before, from .735 to .883. This seems to conform a finding, reported by Kornhauser,⁴ that the average ratings are more reliable than the single ratings.

TABLE II
AVERAGE RELIABILITY COEFFICIENT OF FIVE
TEACHERS WITH RESPECT TO EACH OF
NINETEEN TRAITS AND THEIR
COMPOSITE

Trait	r
Composite -----	.877
Leadership -----	.860
Muscular coordination -----	.860
Self-assertion -----	.859
Sociability -----	.857
Neatness -----	.850
Originality -----	.838
Emotional balance -----	.836
Mental alertness -----	.817
Attentiveness -----	.816
Initiative -----	.811
Intellectual curiosity -----	.810
Accuracy -----	.807
Co-operation -----	.800
Facing reality -----	.788
Persistence -----	.784
Promptness -----	.762
Carefulness -----	.760
Industry -----	.748
Independence -----	.714
Average -----	.812

The coefficient as high as .909, which was found with the average ratings, is quite significant. It indicates that there is slight change, after one semester, in the relative positions of the individuals in the group. In

⁴ A. W. Kornhauser, "Reliability of Average Ratings," *Journal of Personnel Research*, V (1926), 309-17.

other words, one pupil's rating for the second semester can be predicted fairly well from his rating for the first semester. Therefore it may be concluded that the average of the ratings by the five raters is highly reliable and can be safely used for school purposes.

CORRELATION BETWEEN RATINGS
BY TEACHERS

In regard to the correlation between every pair of the five teachers, ten coefficients have been calculated for each of the nineteen traits and for the composite. The average of the

TABLE III
RELIABILITY COEFFICIENTS BETWEEN THE AVERAGE OF FIVE RATINGS AND THE AVERAGE OF FIVE RE-RATINGS WITH RESPECT TO NINETEEN TRAITS AND THEIR COMPOSITE

Trait	r
Accuracy -----	.946
Composite -----	.945
Mental alertness -----	.935
Facing reality -----	.935
Initiative -----	.934
Neatness -----	.933
Leadership -----	.933
Self-assertion -----	.931
Sociability -----	.919
Carefulness -----	.914
Attentiveness -----	.906
Independence -----	.902
Muscular coordination -----	.899
Co-operation -----	.893
Intellectual curiosity -----	.886
Industry -----	.886
Persistence -----	.886
Emotional balance -----	.874
Promptness -----	.860
Originality -----	.856
Average -----	.909

ten coefficients for each trait and the composite is shown in Table IV. The average coefficients range from .292 for independence to .676 for the composite, with an average of .468. The coefficient for the composite is higher than that for any of the nineteen individual traits. That is, the raters agree better in the summation of a number of judgments than in a single judgment.

The coefficients vary greatly with the various traits. The highest coefficient, namely, .627, is more than twice as large as the lowest one, namely, .292. The traits ranking high, with coefficients of .500 and above, are accuracy, neatness, attentiveness, initiative, carefulness, mental alertness, self-assertion, and industry. The traits ranking low, with coefficients below .400, are emotional balance, intellectual curiosity, sociability, facing real-

TABLE IV
AVERAGE CORRELATION COEFFICIENTS BETWEEN EVERY PAIR OF FIVE TEACHERS FOR NINETEEN TRAITS AND THEIR COMPOSITE

Trait	r
Composite -----	.676
Accuracy -----	.627
Neatness -----	.590
Attentiveness -----	.566
Initiative -----	.553
Carefulness -----	.530
Mental alertness -----	.529
Self-assertion -----	.527
Industry -----	.526
Leadership -----	.450
Promptness -----	.449
Co-operation -----	.433
Originality -----	.420
Persistence -----	.401
Muscular coordination -----	.401
Emotional balance -----	.395
Intellectual curiosity -----	.365
Sociability -----	.322
Facing reality -----	.321
Independence -----	.292
Average -----	.468

ity, and independence. The other traits are in the intermediate ranks. This rank order seems to indicate that the teachers' estimations correlate higher on the traits which can be more readily observed. It is to be noted that the specific definitions given to the traits failed to make the teachers' estimations correlate equally well with respect to the various traits. One of the reasons was found to be that the teachers rated not infrequently on different actions listed for one trait. The evidence will be given in the next section.

The average correlation between the teachers, .468, is much lower than the average reliability, .812. And, quite contrary to expectation, the average correlation obtained for these ratings is not so high as the average of the correlations (.550) reported in ten previous studies.⁵ In view of these facts, there must

⁵ James B. Miner, "The Evaluation of a Method for Finely Graded Estimates of Abilities," *Journal of Applied Psychology*, I (1917), 123-33.
A. W. Kornhauser, "Reliability of Average Ratings," *Journal of Personnel Research*, V (1926), 309-17.
G. U. Cleeton and F. B. Knight, "Validity of Character Judgments Based on External Criteria," *Journal of Applied Psychology*, VII (1924), 215-31.
Paul H. Furfey, "An Improved Rating Scale Technique," *Journal of Educational Psychology*, XVII (1926), 45-48.
H. H. Remmers, N. W. Shock, and E. L. Kelly, "An Experimental Study of the Validity of the Spearman-Brown Formula Applied to the Purdue Rating Scale," *ibid.*, XVIII (1927), 197-95.
Edward Webb, *Character and Intelligence*, p. 18. *British Journal of Psychology Monographs*, No. 3 (1915).
H. S. Hayes and D. C. Peterson, "Experimental Development of the Graphic Method," *Psychology Bulletin*, XVIII (1921), 161-71.
J. Slawson, "The Reliability of Judgment of Personal Traits," *Journal of Applied Psychology*, VI (1922), 161-71.
Richard S. Uhrbrock, "Rating Tendencies of Personally Selected Judges," *Journal of Educational Psychology*, XXIII (1932), 594-603.
Conrad, *op. cit.*

be some factors which lowered the correlation between the teachers in the present study. Two possible factors suggest themselves: (1) The lack of completeness of the sub-ratings given by the teachers on the actions listed for many of the traits may have lowered the correlation between their estimations for the traits. (2) the different class-situations in which the teachers observed their pupils may have affected the correlation between their estimations. These two factors will next be examined.

Effect of Lack of Completeness of Sub-ratings. As mentioned before, each of the nineteen traits (except accuracy, carefulness, and persistence) was defined in terms of several actions; and the rating for a trait was the average of the estimates on the actions included. Unfortunately, the teachers frequently omitted, for some reason, to rate the whole group on one or more actions included in a trait. In some cases, one teacher skipped one action, while another skipped another action under the same trait. On account of these omissions the rating by one teacher regarding one trait often included sub-ratings which were different from those included in the rating by another teacher regarding the same trait. Therefore this incompleteness of the sub-ratings would lower the correlation between the teachers to some extent. In order to investigate this effect, one single action under each trait was selected, and the sub-ratings by the teachers on each of these actions were correlated. A comparison of the correlation for each of the single actions with

the corresponding correlation for the trait as a whole will show the effect of the incompleteness of the sub-ratings.

The standard for choosing the actions under each trait was the completeness of the ratings by all the five teachers. The actions chosen are given below:

<i>Trait Under Which the Action Was Listed</i>	<i>Action</i>
Attentiveness	Attends to directions
Mental alertness	Is wide awake and interested
Intellectual curiosity	Asks questions
Originality	In composition
Industry	Works on assigned task without wasting time
Neatness	Neat in work
Promptness	Is on time and ready to work
Initiative	Does more than a minimum amount of work
Independence	Corrects his own work as far as he can
Leadership	Is a leader
Self-assertion	States his opinion with conviction
Facing reality	Accepts just criticism willingly
Co-operation	Enters into group activities voluntarily
Sociability	Is quick to make friendly approaches
Emotional balance	Appears calm
Muscular coordination	In writing

The average correlation coefficient between every pair of the five teachers for each of these actions is presented in Table V. The coefficients are, on the average, .376. This average is lower than for the heading traits, which is .446.

TABLE V
CORRELATION COEFFICIENTS BETWEEN TEACHERS FOR SIXTEEN SINGLE ACTIONS COMPARED WITH CORRELATION COEFFICIENTS FOR THEIR HEADING TRAITS

Trait	r for Single Action	r for Trait	Difference
Attentiveness.....	.229	.566	— .337
Mental alertness.....	.376	.529	— .153
Intellectual curiosity.....	.516	.365	.151
Originality.....	.520	.420	.100
Industry.....	.315	.526	— .211
Neatness.....	.372	.590	— .218
Promptness.....	.315	.449	— .134
Initiative.....	.507	.553	— .064
Independence.....	.163	.292	— .129
Leadership.....	.483	.350	.033
Self-assertion.....	.552	.527	.025
Facing reality.....	.138	.321	— .183
Co-operation.....	.326	.433	— .017
Sociability.....	.436	.320	.116
Emotional balance.....	.454	.395	.059
Muscular coordination.....	.408	.401	.007
Average.....	.376	.446	— .064

The comparison of the coefficient for each of the single actions with the coefficient for its heading trait under which the action was included is also given in Table V. Seven of the sixteen coefficients show an increase of from .007 for the action of muscular coordination to .151 for the action of intellectual curiosity. The intermediate increases are for the action of originality, that of leadership, of self-assertion, of sociability, and of emotional balance. Nine coefficients show a decrease of from .025 for the action of self-assertion to .337 for the action of attentiveness. The intermediate decreases are for the action of mental alertness, that of industry, of neatness, of promptness, of initiative, of independence, of facing reality, and of cooperation.

The increases of correlation between the teachers regarding the single actions are significant, although they are small; while the decreases are to be expected. As discussed before, different raters correlate higher when a composite of several ratings is used than they do for a single rating. According to this theory, each single action should yield a lower correlation between teachers than its heading trait, since the latter is a composite of several single actions. On the contrary, the

coefficients for the single actions are larger, in almost half of the cases, than the coefficients for their heading traits; and the average of the former coefficients, namely, .376, is close to that of the latter coefficients, namely, .446. This result may be taken as evidence for concluding that the incompleteness of the sub-ratings under some of the traits lowered the correlation between the teachers to a certain extent.

The neglect of estimating on some of the actions listed under a trait seems to indicate that raters tend to differ in what to observe or judge upon in a trait rating.

Effects on Judgment of Different Class-Situations. According to Hanna, different class-situations will cause teachers to judge differently on personality traits.⁶ In order to ascertain this effect on the ratings of the present study, the teachers were matched in three groups according to certain aspects of resemblance; and the correlation coefficients between the ratings by the teachers within each group are to be compared.

The first is a grade group which was composed of one pair of raters teaching fifth

⁶ Joseph V. Hanna, "Variable Factors Encountered in the Rating of Students," *School Science and Mathematics*, XXV (1925), 481-83.

TABLE VI

CORRELATION COEFFICIENTS BETWEEN TEACHERS MATCHED ACCORDING TO CERTAIN ASPECTS OF RESEMBLANCE, FOR NINETEEN TRAITS AND THEIR COMPOSITE

Trait	Teachers of Same Grade		Teachers of Same Subject		Teachers of Different Subjects in Different Grades	
	r	Rank	r	Rank	r	Rank
Attentiveness.....	.561	6	.516	8	.596	3
Mental alertness.....	.567	5	.471	12	.521	7
Intellectual curiosity.....	.338	19	.395	14	.390	16
Originality.....	.403	14	.518	6.5	.396	14
Industry.....	.534	8	.575	4	.497	9
Persistence.....	.433	12	.311	18	.413	13
Accuracy.....	.571	4	.683	1	.655	2
Carefulness.....	.499	9	.494	10	.579	6
Neatness.....	.590	2	.605	3	.583	4
Promptness.....	.485	10	.329	17	.432	11
Initiative.....	.556	7	.493	11	.582	5
Independence.....	.320	20	.347	15	.238	20
Leadership.....	.421	13	.518	6.5	.459	12
Self-assertion.....	.573	3	.506	9	.484	9
Facing reality.....	.371	18	.236	19	.391	15
Co-operation.....	.377	17	.448	13	.481	10
Sociability.....	.389	16	.172	20	.324	19
Emotional balance.....	.442	11	.345	16	.375	17
Muscular coordination.....	.392	15	.528	5	.349	18
Composite.....	.700	1	.651	2	.663	1
Average.....	.474	--	.458	--	.472	--

grade and three pairs of raters teaching sixth grade. The second is a subject group which was made up with one pair of raters who taught English and arithmetic and another pair who taught social studies. The last or a heterogeneous group included four pairs of raters, each pair teaching two different subjects in two different grades. Thus there resulted four coefficients from the first group, two from the second group, and four from the third group.

If the subject which a teacher taught influenced her attitude most in judging the personality traits of the pupils, the correlation between the teacher of the subject group would be higher than that between the teachers of either of the other two groups. If the grade in which a teacher taught influenced her attitude most in judging the personality traits of the pupils, the correlation between the teachers of the grade group would be higher than that between the teachers of either of the other two groups. Lastly, if both the subject which a teacher taught and the grade in which she taught affected markedly her attitude in rating the pupils, the correlation obtained for the subject group and that for the grade group would both be higher than the correlation obtained for the heterogeneous group, composed

of teachers who taught different subjects in different grades.

The average correlation coefficient for each group with respect to each trait is presented in Table VI. The three average coefficients, namely, .474, .458, and .472, are almost equal. Moreover, the rank order of the traits according to their coefficients as shown in Table VI varies only slightly with the different groups. Thus it is obvious that the subjects which the teachers taught and the grade in which they taught did not particularly influence their observation and estimation of the personality traits of the pupils.

This result is not in harmony with Hanna's finding. The explanation is probably that the school subjects the teachers taught, namely, English and arithmetic, social studies, and French, are rather similar regarding activities of students, so that they have no different effects on the teachers' observation of the pupils' personality traits.

When the teachers of different subjects and of different grades rate in a similar way, their estimates on a trait will indicate the same personality quality. The ratings by the different teachers can then be combined to obtain a higher reliability for practical purposes than that of one teacher.

TABLE VII

"INTERNAL" VALIDITY COEFFICIENTS OF THE RATINGS BY EACH OF FIVE TEACHERS FOR SIXTEEN TRAITS* AND THE COMPOSITE

Trait	Teacher				
	A	B	C	D	E
Attentiveness.....	.580	.723	.655	.783	.712
Mental alertness.....	.651	.647	.652	.661	.690
Industry.....	.679	.567	.572	.817	.784
Persistence.....	.436	.426	.558	.211	.682
Accuracy.....	.742	.703	.702	.858	.691
Carefulness.....	.593	.498	.800	.699	.730
Neatness.....	.713	.662	.676	.786	.711
Promptness.....	.580	.433	.633	.684	.618
Initiative.....	.663	.531	.744	.715	.750
Independence.....	.414	.523	.430	.537	.248
Self-assertion.....	.659	.675	.716	.495	.731
Facing reality.....	.258	.438	.621	.606	.559
Co-operation.....	.430	.667	.532	.631	.607
Sociability.....	.306	.630	.603	.565	.242
Emotional balance.....	.494	.481	.567	.483	.650
Muscular coordination.....	.592	.437	.453	.596	.648
Composite.....	.746	.744	.766	.824	.804
Average.....	.561	.576	.628	.644	.639

*Intellectual curiosity, originality, and leadership to not appear in the table, because the data on them are incomplete.

MAJORITY JUDGMENT AS THE CRITERION FOR VALIDATION

The correlation between the judgment of each of the five teachers and the consensus of the other four may be regarded as the coefficient of "internal" validity for each teacher. Since this criterion is based on internal consistency, the resulting coefficient does not tell whether the estimations are valid with reference to an outside criterion.

The coefficients of "internal" validity are presented in Table VII. Three traits, namely, intellectual curiosity, originality, and leadership, have been omitted from this section, because the data on them were not quite complete for the purposes of this part of the investigation. The average coefficients of "internal" validity range from .561 for teacher A to .644 for teacher D. This range of variation among them, namely, .083, is quite small and insignificant. Thus the ratings by the five raters are equally valid as judged from the internal consistency criterion. The average validity of the five raters, .610, indicates that, in general, one teacher's judgment is rather close to the combined judgments of the other four.

With the average validity .610 and the average reliability .812 (Table II) for one teacher, the average of the judgments of the five teachers will yield a validity coefficient of .739 according to Holzinger's formula 51.⁷ This figure is quite high. Theoretically only the average of ratings by an infinite number of raters can show perfect validity; but it is far from a practical possibility. The coefficient of .739 may be interpreted to mean that the average of the ratings by the five teachers is fairly close to the true rating.

SUMMARY

1. The reliability coefficients for the five teachers, as measured by the average of the correlations between ratings and re-ratings with respect to the nineteen traits and their composite, varied from .735 for teacher B to .883 for teacher C. This range of variation being rather small, the five teachers may be regarded as equally reliable in judgment.

The average of the average reliability coefficients of the five teachers was .812. It is much higher than .650, the average of the reliability coefficients reported in previous studies. The favorable conditions under which

the ratings were made appear to have raised the reliability.

2. When the average of the five ratings for the first semester was correlated with that of the five re-ratings for the second semester, the reliability coefficients ranged from .856 for originality to .946 for accuracy, with an average of .909. This coefficient is quite high. It indicates that the average rating can be safely used for educational guidance.

3. The average correlation coefficients between the ratings by every pair of the five teachers on the various traits and their composite ranged from .292 for independence to .676 for the composite, with average of .468. The wide range of variation of the correlations for various traits indicates that the teachers agreed in different degrees in judging the various traits. The higher correlations appear to come with the traits which are more accessible to observation.

The disparity between the average correlation, .468, and the average reliability, .812, is quite large. One of the factors causing the correlation between the ratings of different teachers to be too low was found to be the incompleteness of the sub-ratings on many of the traits.

The different class-situations in which the teachers observed the pupils affected the teachers' estimations little, if any.

4. The "internal" validity of the rating by a single teacher, on the average, was .610, which is fair. The average ratings by five teachers would yield a higher coefficient, namely, .739. This coefficient may be interpreted to mean that the average rating of each pupil is reasonably close to his true rating.

5. On the whole, the estimations of the five teachers were quite reliable; but the correlations between them were not so high as might be expected. The disparity of the two phases of consistency seems to be due partially to the fact that raters differ in what they observe.

PART II

HALO EFFECT

MEANING OF HALO EFFECT

By halo effect is meant a systematic error of judgment which is usually found in ratings. This error is due to the fact that the rater's general impression of a person rated affects

⁷ *Op. cit.*, p. 170.

his ratings with respect to specific traits; in other words, a rater can hardly judge a person with respect to specific traits independently of his general impression of the person being rated. The name of this error was first given in 1920 by Thorndike.⁸ Nevertheless, the error was discovered by Wells as early as 1907 in his statistical study of literary merit.⁹ In that study he had a graduate English class rate ten authors on ten literary qualities. Commenting on the halo effect, he says:

There is noted introspectively a tendency to grade for general merit at the same time as for the qualities, and to allow an individual's general position to influence his position in the qualities. This would be the case especially in the case of those qualities that were ill-defined in the minds of the subjects and tended to be interpreted rather in terms of general merit . . . This would make the correspondence of such qualities appear closer than they are.

Webb pointed out the same error in his study of intelligence rating. He says, "The observers in estimating intelligence qualities are biased in the direction of marking subjects who possess other desirable qualities too highly and vice versa."¹⁰ Thorndike reported, from a study by Knight, the correlations between general merit as a teacher and forty-five traits.¹¹ Most of them were between .50 and .70. He concluded that these correlations were too high and too even, and that the ratings on the specific traits of a teacher must have been influenced by the rater's general impression of her. The same error was also found by Rugg in his study of the Army Rating Scale.¹² He says, "We judge our fellows in terms of a general attitude toward them; and there is dominating this mental attitude toward particular qualities." Thus, it seems quite certain that halo is a reality in ratings, and that it is a troublesome factor and tends to make ratings invalid.

So far, however, there has not been published any study to investigate the size of the halo except the short paper of Symonds in 1925.¹³ In that study, he had a group of

pupils rated by two judges with respect to seven traits, which were honesty, obedience, courtesy, orderliness, cleanliness, sportsmanship, and promptness. He took the average of the ratings of each pupil with respect to the seven traits as the general impression. After he had determined the correlation coefficients between the two raters for each of the seven traits and the coefficient for the general impression, he used the technique of partial correlation to eliminate the general impression from each of the seven correlations for the individual traits. Then for each trait he took the difference between the raw and the partial correlations as the size of the halo effect in the rating on that trait. As a result, six of the seven correlations were found to have been raised by the halo by amounts ranging from .086 to .423, while one was lowered by .083. The halo effect in the seven correlations was, on the average, .245.

Although Symonds' study is suggestive, his method is not justified for two reasons. In the first place, the general impression which a rater had of each of the pupils might not be the average of his seven traits; it might quite possibly be something outside of them. Secondly, his assumption that halo tends to raise the correlations between the ratings by two raters on individual traits is not in harmony with its definition. According to the foregoing discussion, halo tends only to raise the intercorrelations of the ratings on different traits by one rater. Consequently, what he found cannot be properly regarded as the halo effect.

METHOD OF INVESTIGATION

According to the foregoing discussion, halo is the effect upon ratings of a general attitude of a rater toward a subject. This general attitude affects his attitudes concerning the various traits. Because of this effect the correlation between two traits, according to the ratings by one rater, tends to be higher than it should be. On the other hand, since two raters are not likely to take the same attitude or to be under the same prejudice toward an individual rated, the correlation between two traits, according to the ratings by two different raters, would be relatively free from the halo effect. Hence the difference between the former and the latter correlation coefficients may be regarded as the halo effect contained in the ratings by one rater.

⁸ E. L. Thorndike, "A Constant Error in Psychological Rating," *Journal of Applied Psychology*, IV (1920), 25-29.

⁹ F. L. Wells, *A Statistical Study of Literary Merit*, *Archives of Psychology*, No. 7 (1909).

¹⁰ *Op. cit.*, p. 70. ¹¹ *Op. cit.*

¹² H. O. Rugg, "Is the Rating of Human Character Practicable?" *Journal of Educational Psychology*, XIII (1922), 30-42.

¹³ "Notes on Rating," *Journal of Applied Psychology*, IX (1925), 188-95.

Theoretically, the correlation between two traits, according to the ratings by one rater, would be unity when the halo reaches its maximum; as the larger the halo, the less the differentiation among the different traits of an individual. And it would be zero when there is no halo effect, provided that the traits are really independent of each other and that they do not overlap in the rater's conception of them.

In order not to make this investigation too laborious to handle, three raters only have been included. They are A, B, and C, the three teachers of the sixth grade.

The intercorrelations of the ratings on the nineteen traits by each of the three teachers, A, B, and C, are indicated by $r_{AA'}$, $r_{BB'}$, and $r_{CC'}$, respectively. All these intercorrelations are supposed to contain the halo effect. The intercorrelations of the ratings by every possible pair of the three teachers are indicated by r_{AB} , r_{BA} , r_{AC} , r_{CA} , r_{BC} and r_{CB} . All of these contain no halo. Each coefficient r_{AB} is defined as the correlation between the ratings of two traits, the first of which was rated by teacher A, the second by teacher B. Each coefficient r_{BA} is the correlation between the ratings of two traits, the first of which was rated by teacher B, the second by teacher A. In like manner are r_{AC} distinguished from r_{CA} , and r_{BC} from r_{CB} . The averages of the sets of six intercorrelations are defined as the "true" intercorrelations, in the sense that they are free from halo effect, and that, as resulting from the judgments of three teachers, they contain few errors of individual judgment.

The method of determining the halo may be formulated as follows:

Let r_T = the "true" intercorrelations among the nineteen traits,

r_H = the average of $r_{AA'}$, $r_{BB'}$, and $r_{CC'}$,

h = the size of the halo contained in r_H , or the average halo;

then $r_T = \frac{r_{AB} + r_{BA} + r_{AC} + r_{CA} + r_{BC} + r_{CB}}{6} \dots (1)$

$r_H = \frac{r_{AA'} + r_{BB'} + r_{CC'}}{3}, \dots \dots \dots (2)$

$h = r_H - r_T; \dots \dots \dots (3)$

and, $r_{AA'} - r_T$ = the halo contained in $r_{AA'}$, (4)

$r_{BB'} - r_T$ = the halo contained in $r_{BB'}$, (5)

$r_{CC'} - r_T$ = the halo contained in $r_{CC'}$, (6)

DETERMINATION OF HALO

The intercorrelations of the nineteen ratings by each of the three teachers, or $r_{AA'}$, $r_{BB'}$, and $r_{CC'}$, are not presented here because of limited space. Since the three sets of in-

tercorrelations are all assumed to contain halo, the highest among the three would be from the rater whose estimations were most affected by halo, assuming that the intercorrelations would otherwise be equal. The average of the coefficients $r_{CC'}$ is .591, the average of $r_{BB'}$ is .380, and the average of $r_{AA'}$ is .389. Among the three teachers, apparently, C's estimations contain the largest halo; while A's and B's are roughly on a par in this respect. The difference between the average $r_{CC'}$ and $r_{AA'}$ is .202; and the difference between the average $r_{CC'}$ and $r_{BB'}$ is .211. These differences appear quite significant, although the means of comparison is rather crude.

The comparison of two correlation coefficients with their probable errors does not apply here, because each of the three averages is obtained from a group of 171 coefficients. The probable error of an average of a number of coefficients should be smaller than that of a single coefficient. Unfortunately the formula for its computation is not available.

It is rather safe to say that the ratings by C contain a larger amount of halo than those by either A or B; the two latter teachers differ only slightly in this respect.

It should be noted that the variation among the teachers in the average size of the intercorrelations could be affected by their variation either in reliability or in validity. In other words, the reliable and valid rater may tend to obtain a higher correlation in estimating two traits than does an unreliable and invalid rater. This effect, however, does not occur in the present case, for the three teachers of the present study are similar, as discussed in Part I of this paper, both in reliability and in validity.

In view of the fact that the intercorrelations for one rater may contain errors of judgment other than halo, it is more justifiable for the determination of the halo to use r_H , the average of the three sets of intercorrelations by the three teachers separately, since the average intercorrelations may be considered as near a typical set of intercorrelations for one teacher and they contain few errors of individual judgment besides halo. The result of this comparison will show the amount of halo which usually operates in rating.

The actual amount of the halo contained in r_H cannot be obtained until the true intercorrelations for the various traits are available; since the determination of the halo involves both of the two sets of intercorrela-

tions. The true intercorrelations should be the average of an infinite number of sets of intercorrelations for the nineteen traits, each pair of which should be estimated by every pair of an infinite number of raters. This is, of course, an actual impossibility. In the present study, only an approximation to the true intercorrelations has been obtained, that is, r_T by formula (1).

According to formula (3), the differences, r_H minus r_T , for each pair of the traits, give the amount of the halo. This comparison is presented in Table VIII.

The resulting differences for the various combinations of the nineteen traits are practically all positive. The differences range from $-.064$ to $.385$, with an average of $.140$; and over two thirds of them are larger than $.100$. The consistent appearance of positive differences may be taken as evidence for the existence of halo effect in the ratings.

It seems reasonable that the ratings on the various traits were affected by the halo in different degrees; as the ratings on the more specific traits naturally would be less affected and the ratings on the more general traits would be more affected by the halo.

TABLE VIII

DIFFERENCES BETWEEN AVERAGES OF INTERCORRELATIONS FOR THREE TEACHERS (r_H),
AND "TRUE" INTERCORRELATIONS (r_T)

Pair of Traits		r_H	r_T	$r_H - r_T$
Mental alertness v.	industry	.453	.325	.128
" "	carefulness	.483	.358	.125
" "	initiative	.732	.494	.238
" "	neatness	.288	.244	.044
" "	sociability	.479	.262	.217
" "	emotional balance	.588	.349	.239
" "	muscular coordination	.270	.206	.064
" "	attentiveness	.489	.271	.218
" "	intellectual curiosity	.638	.441	.197
" "	persistence	.616	.355	.261
" "	originality	.453	.354	.099
" "	accuracy	.678	.473	.205
" "	promptness	.607	.374	.233
" "	independence	.566	.388	.178
" "	leadership	.582	.456	.126
" "	self-assertion	.810	.530	.280
" "	cooperation	.350	.243	.107
" "	facing reality	.588	.363	.225
Industry v.	carefulness	.607	.402	.205
" "	initiative	.646	.460	.186
" "	neatness	.447	.352	.095
" "	sociability	.168	.067	.101
" "	emotional balance	.515	.313	.202
" "	muscular coordination	.287	.271	.016
" "	attentiveness	.836	.413	.423
" "	intellectual curiosity	.341	.277	.064
" "	persistence	.759	.374	.385
" "	originality	.440	.331	.109
" "	accuracy	.707	.420	.287
" "	promptness	.672	.396	.276
" "	independence	.600	.352	.248
" "	leadership	.284	.226	.058
" "	self-assertion	.326	.222	.104
" "	cooperation	.586	.314	.272
" "	facing reality	.599	.371	.228
Carefulness v.	initiative	.480	.425	.055
" "	neatness	.682	.431	.251
" "	sociability	.163	.164	.001
" "	emotional balance	.499	.298	.201
" "	muscular coordination	.455	.329	.126
" "	attentiveness	.692	.339	.353
" "	intellectual curiosity	.334	.280	.054
" "	persistence	.528	.319	.209

TABLE VIII (Continued)
DIFFERENCES BETWEEN AVERAGES OF INTERCORRELATIONS FOR THREE TEACHERS (r_H),
AND "TRUE" INTERCORRELATIONS (r_T)

Pair of Traits		r_H	r_T	$r_H - r_T$
Carefulness v.	originality	.369	.310	.059
"	" accuracy	.653	.383	.270
"	" promptness	.526	.346	.180
"	" independence	.610	.335	.255
"	" leadership	.321	.237	.084
"	" self-assertion	.331	.210	.121
"	" cooperation	.376	.275	.101
"	" facing reality	.500	.305	.195
Initiative v.	neatness	.395	.385	.010
"	" sociability	.373	.231	.142
"	" emotional balance	.549	.400	.149
"	" muscular coordination	.335	.291	.044
"	" attentiveness	.565	.395	.170
"	" intellectual curiosity	.605	.425	.180
"	" persistence	.676	.457	.219
"	" originality	.584	.394	.190
"	" accuracy	.729	.495	.234
"	" promptness	.607	.457	.150
"	" independence	.608	.399	.209
"	" leadership	.506	.407	.099
"	" self-assertion	.646	.429	.217
"	" cooperation	.472	.359	.113
"	" facing reality	.666	.420	.246
Neatness v.	sociability	.104	.057	.047
"	" emotional balance	.286	.218	.068
"	" muscular coordination	.549	.354	.195
"	" attentiveness	.452	.252	.200
"	" intellectual curiosity	.189	.172	.017
"	" persistence	.401	.296	.105
"	" originality	.278	.192	.086
"	" accuracy	.412	.307	.105
"	" promptness	.395	.298	.097
"	" independence	.267	.260	.007
"	" leadership	.232	.194	.038
"	" self-assertion	.121	.117	.004
"	" cooperation	.314	.225	.089
"	" facing reality	.304	.223	.081
Sociability v.	emotional balance	.426	.337	.089
"	" muscular coordination	.144	.178	— .034
"	" attentiveness	.215	.081	.134
"	" intellectual curiosity	.255	.285	— .030
"	" persistence	.272	.182	.090
"	" originality	.328	.220	.108
"	" accuracy	.216	.150	.066
"	" promptness	.281	.224	.057
"	" independence	.269	.234	.035
"	" leadership	.484	.303	.181
"	" self-assertion	.559	.373	.186
"	" cooperation	.512	.229	.283
"	" facing reality	.409	.209	.200
Emotional balance v.	muscular coordination	.273	.252	.021
"	" attentiveness	.421	.303	.118
"	" intellectual curiosity	.556	.313	.243
"	" persistence	.365	.315	.050
"	" originality	.419	.317	.102
"	" accuracy	.507	.324	.183
"	" promptness	.480	.287	.193
"	" independence	.523	.316	.207
"	" leadership	.448	.282	.116
"	" self-assertion	.529	.339	.190
"	" cooperation	.542	.318	.224
"	" facing reality	.567	.312	.255

TABLE VIII (Continued)

DIFFERENCES BETWEEN AVERAGES OF INTERCORRELATIONS FOR THREE TEACHERS (r_H),
AND "TRUE" INTERCORRELATIONS (r_T)

Pair of Traits		r_H	r_T	$r_H - r_T$
Muscular coordination v.	attentiveness	.295	.206	.089
"	" intellectual curiosity	.261	.237	.024
"	" persistence	.202	.230	— .028
"	" originality	.367	.276	.091
"	" accuracy	.360	.295	.065
"	" promptness	.217	.248	— .031
"	" independence	.273	.282	— .009
"	" leadership	.249	.209	.040
"	" self-assertion	.183	.230	— .047
"	" cooperation	.284	.227	.057
"	" facing reality	.157	.221	— .064
Attentiveness v.	intellectual curiosity	.321	.259	.061
"	" persistence	.574	.315	.259
"	" originality	.456	.382	.074
"	" accuracy	.623	.390	.233
"	" promptness	.633	.332	.301
"	" independence	.577	.399	.178
"	" leadership	.296	.199	.097
"	" self-assertion	.299	.168	.131
"	" cooperation	.547	.301	.246
"	" facing reality	.523	.320	.203
Intellectual curiosity v.	persistence	.445	.315	.130
"	" originality	.462	.336	.126
"	" accuracy	.558	.346	.212
"	" promptness	.371	.320	.051
"	" independence	.427	.306	.121
"	" leadership	.465	.376	.089
"	" self-assertion	.665	.455	.210
"	" cooperation	.178	.214	— .036
"	" facing reality	.430	.303	.127
Persistence v.	originality	.394	.326	.068
"	" accuracy	.543	.391	.152
"	" promptness	.727	.365	.362
"	" independence	.594	.323	.271
"	" leadership	.340	.265	.075
"	" self-assertion	.474	.305	.170
"	" cooperation	.494	.272	.222
"	" facing reality	.694	.336	.358
Originality v.	accuracy	.525	.379	.146
"	" promptness	.305	.343	— .038
"	" independence	.383	.355	.028
"	" leadership	.593	.281	.312
"	" self-assertion	.449	.329	.120
"	" cooperation	.351	.230	.121
"	" facing reality	.386	.299	.087
Accuracy v.	promptness	.506	.362	.144
"	" independence	.662	.410	.252
"	" leadership	.463	.339	.124
"	" self-assertion	.561	.364	.197
"	" cooperation	.393	.198	.195
"	" facing reality	.597	.336	.261
Promptness v.	independence	.570	.288	.282
"	" leadership	.357	.323	.034
"	" self-assertion	.425	.306	.119
"	" cooperation	.525	.302	.223
"	" facing reality	.622	.362	.260
Independence v.	leadership	.361	.273	.088
"	" self-assertion	.437	.308	.129
"	" cooperation	.450	.326	.124
"	" facing reality	.513	.307	.206

TABLE VIII (Continued)
DIFFERENCES BETWEEN AVERAGES OF INTERCORRELATIONS FOR THREE TEACHERS (r_H),
AND "TRUE" INTERCORRELATIONS (r_T)

Pair of Traits	r_H	r_T	$r_H - r_T$
Leadership v. self-assertion.....	.653	.480	.173
" " cooperation.....	.335	.232	.103
" " facing reality.....	.400	.214	.186
Self-assertion v. cooperation.....	.310	.212	.098
" " facing reality.....	.495	.303	.192
Cooperation v. facing reality.....	.545	.278	.267
Average.....	.456	.306	.140

The average of the correlations due to the halo is .140, which may be regarded as the average amount of the halo found in the ratings. It is about half as large as the average of the "true" intercorrelations, namely, .306. Thus the halo effect in the ratings is quite significant. A rating on one trait does not have a completely specific value, when it contains halo. Its validity is thus impaired.

SUMMARY

1. According to the evidence obtained, the existence of the halo effect in the ratings has been proved. Although the ratings were given with respect to specific actions in terms of which the traits were defined, the halo was still a significant factor operative in the teachers' estimations.

2. Raters probably differ to some extent in making this error of judgment, and traits for estimation are subject to the halo effect in different degrees. From the present data, it is .140 on the average. The validity of the ratings is thus impaired.

PART III
FACTOR ANALYSIS

The significance of personality ratings may be examined with the technique of factor analysis. In the ordinary sense, a trait is supposedly a mental entity; whence the ratings on the various traits should be independent of one another. In human judgment, however, various traits can hardly be independently rated, because irrelevant as well as relevant factors often operate in the rater's mind. What are the operative factors? And what proportion does each of them play in accounting for the variance of the ratings? These may be answered through factor analysis.

This study is based on ratings by three teachers of their 100 pupils on nineteen traits.

The source of the data was described in Part I.

The analysis was begun with the "true" intercorrelations, or r_T , for the nineteen traits. As discussed in Part II, the "true" intercorrelations are approximately free from halo effect.

For analysis, the B-factor technique¹⁴ was applied. According to the uniformity of the B-coefficients, there seems to be only one general factor running through this table of correlations. After the removal of this general factor, the residual correlations appear altogether insignificant. The distribution of the residuals with its mean, standard deviation and probable error are given below. The probable error of zero correlation is also given for comparison.

TABLE IX
FREQUENCY OF RESIDUAL CORRELATIONS

Residuals	Frequency
.1700-.1899	2
.1500-.1699	2
.1300-.1499	1
.1100-.1299	4
.0900-.1099	4
.0700-.0899	5
.0500-.0699	7
.0300-.0499	15
.0100-.0299	23
-.0100-.0099	33
-.0300-.0101	33
-.0500-.0301	17
-.0700-.0501	10
-.0900-.0701	5
-.1100-.0901	7
-.1300-.1101	2
-.1500-.1301	—
-.1700-.1501	2
Number of residuals	171
Mean	.00003
S. D.	.0583
P. E.	.0393
P. E. of zero correlation	.0674

¹⁴ Preliminary Report on Spearman-Holzsinger Unitary Trait Study, No. 7 K. J. Holzinger and F. Swineford, *An Objective Method for Allocating Tests for the Bi-Factor Pattern and Steps for Pattern Evaluation*, Prepared at the Statistical Laboratory, Department of Education, The University of Chicago, 1936. (Lithoprinted.)

This distribution is nearly normal, with a mean of about zero; and even the largest of the residuals is less than three times the probable error of zero correlation. Thus, all of the residual correlations may be regarded as due to chance. The application of the tetrad criterion to the comparatively large residuals also failed to reveal any extra factors. The evidence seems sufficient for concluding that only one general factor runs through the "true" intercorrelations.

The nature of this general factor may be examined from its loadings in the various traits; they are presented in rank order as follows:

TABLE X
GENERAL FACTOR LOADINGS

Ratings	Loadings of General Factor
Initiative -----	.7493
Mental alertness -----	.6568
Accuracy -----	.6463
Promptness -----	.5979
Independence -----	.5924
Industry -----	.5920
Carefulness -----	.5797
Persistence -----	.5772
Self-assertion -----	.5696
Originality -----	.5680
Intellectual curiosity -----	.5623
Emotional balance -----	.5618
Facing reality -----	.5497
Leadership -----	.5384
Attentiveness -----	.4810
Cooperation -----	.4732
Muscular coordination -----	.4509
Neatness -----	.3715
Sociability -----	.3715

The loadings of this factor in the various ratings are quite high. In view of the fact that the traits of the present study are all desirable qualities, the general factor obtained from them indicates probably some good personality quality which is fundamental in nature.

If a rater has only a vague conception of the traits, an overlapping may result in his various ratings. This overlapping would raise the intercorrelations between the traits to too high a value, and thereby create a false general factor. This error is probably very small in the present case, if it occurs at all. The reason is that the traits were defined in terms of specific actions, which could hardly be misconceived.

One might again suspect that the general factor is due to age, which is usually responsible for positive correlations between mental

abilities or traits. This explanation does not apply to any great extent in the present data, since the subjects of the present ratings were in the same grades and were of about the same age level.

All these facts seem to indicate that the general factor found from these data is some fundamental quality, or at least that it is so according to the judgments of the teachers.

Further analysis of the loadings of the general factor in the various ratings will clarify the nature of this factor. As shown in the foregoing table, the loadings range from .3715 to .7493. Initiative has the highest loading. Next rank mental alertness, accuracy, promptness, independence, industry, carefulness, persistence, and self-assertion. All these traits ranking above the median loading are somewhat volitional in nature and are essential for school achievement. The two intellectual traits, originality and intellectual curiosity, rank in the middle. The emotional traits, emotional balance and facing reality, come below the middle. Then come leadership and attentiveness. Down at the bottom of the rank order are the social and physical traits, namely, cooperation, muscular coordination, neatness, and sociability. This order of the loadings seems to indicate that the general factor is in close relation to the volition to achieve in school work. This volition being contrary to biological drives is probably a result of training. Since most of the loadings in the various ratings are quite high, with a median of .5680, this factor is probably pervasive in scope; it may run also through many other traits. It is a fact that the volition to achieve in a task is usually responsible for developing many of the desirable traits for that task.

The volition to achieve may be called a general factor of will-power. It was also found by Webb¹⁵ and advocated by Spearman¹⁶ as one general factor underlying human personality and character.

In view of the fact that the estimates may be different from measurements on performance tests, it is an interesting question to ask whether the general factor would be as large as is found from the ratings, or if it would come out at all, if the traits were measured with performance tests. Its answer is beyond the scope of this study.

¹⁵ E. Webb, *Character and Intelligence*, *British Journal of Psychology Monographs*, No. 3 (1915), p. 60.

¹⁶ C. Spearman, *Abilities of Man* (New York: Macmillan Co., 1927), p. 347.

From the intercorrelations due to the halo effect, ($r_H - r_T$), a general factor has also been obtained by the same procedure. The residuals after the removal of this factor are practically all within the range of three times the probable error of zero correlation. As this set of correlations is presumably due to the halo effect, as discussed in Part II, the factor obtained therefrom is the general factor of halo. Its loadings in the various ratings are as follows:

TABLE XI
HALO FACTOR LOADINGS

Ratings	Loadings of Halo
Persistence -----	.7116
Facing reality -----	.5309
Attentiveness -----	.5263
Accuracy -----	.5025
Mental alertness -----	.4797
Emotional balance -----	.4323
Promptness -----	.4271
Initiative -----	.4244
Carefulness -----	.4214
Cooperation -----	.4163
Independence -----	.4157
Self-assertion -----	.3829
Industry -----	.3048
Leadership -----	.3016
Neatness -----	.2847
Originality -----	.2729
Sociability -----	.2684
Intellectual curiosity -----	.2646
Muscular coordination -----	.0768

The halo factor runs through the rating on every one of the traits; its loadings vary

widely with the traits, ranging from .0768 to .7116. As is to be expected, the rating on the trait which can be most readily observed, namely muscular coordination, has almost zero loading with the halo factor. Similarly, the ratings on neatness, intellectual curiosity, originality, and sociability have also comparatively low saturations of the halo. On the other hand, the ratings on the somewhat vague traits, such as persistence, facing reality, attentiveness, accuracy, mental alertness, emotional balance, etc. have comparatively high loadings of the halo.

The combination of the loadings of the halo factor and the general factor will produce approximately r_H or the average intercorrelations by the three raters separately; they were shown in Table I in Part II.

The uniqueness of each rating can be determined after the coefficients of the common factors are known. If C designates the coefficient of a factor, H and W designate the halo factor and the general factor of will-power respectively, and U designates the uniqueness; the coefficient of the latter may be obtained from the formula, $C^2_H + C^2_W + C^2_U = 1$. The uniqueness of each rating contains both the specific factor and the chance error factor. The latter is the unreliability of each rating, and its coefficient is $\sqrt{1 - r^2_{I_1}}$. The coefficient of each specific factor can be obtained by removing the chance factor from the uniqueness by the formula, $C^2_s + C^2_e =$

TABLE XII
FACTOR PATTERN

Traits	Factors with Coefficients				Total Variance
	H	W	S	E	
1 Mental alertness -----	.4797	.6568	.5482	.1949	1.0001
2 Industry -----	.3048	.5920	.6787	.3100	1.0001
3 Carefulness -----	.4214	.5797	.6242	.3110	.9999
4 Initiative -----	.4244	.7493	.4677	.1990	.9999
5 Neatness -----	.2847	.4416	.8378	.1482	.9999
6 Sociability -----	.2684	.3715	.8793	.1300	1.0000
7 Emotional Balance -----	.4323	.5618	.6795	.1890	.9999
8 Muscular coordination -----	.0768	.4509	.8748	.1595	.9999
9 Attentiveness -----	.5263	.4810	.6655	.2210	1.0000
10 Intellectual curiosity -----	.2646	.5623	.7698	.1460	1.0000
11 Persistence -----	.7116	.5772	.3074	.2570	1.0001
12 Originality -----	.2729	.5680	.7597	.1610	1.0001
13 Accuracy -----	.5025	.6463	.5271	.2280	1.0000
14 Promptness -----	.4271	.5979	.6265	.2600	1.0000
15 Independence -----	.4157	.5924	.6077	.3265	1.0000
16 Leadership -----	.3016	.5384	.7758	.1320	1.0001
17 Self-assertion -----	.3829	.5696	.7118	.1490	.9999
18 Cooperation -----	.4163	.4732	.7452	.2180	1.0000
19 Facing reality -----	.5309	.5497	.5986	.2400	.9999
Total Variance -----	3.2549	6.0012	8.8405	.9034	19.0000
Per Cent Variance -----	17.13	31.59	46.53	4.75	100.

C^2_U .¹⁷ The resulting factor pattern is presented in Table XII.

The coefficients of the specific factors vary widely with the various ratings, with a range from .3074 to .8793. The ratings on muscular coordination, neatness, and sociability have specific loadings above .8000; while those on persistence, initiative, mental alertness, accuracy, and facing reality have comparatively low specific loadings, below .6000. The former traits do seem more accessible to observation than the latter. So the result is quite sensible.

The chance error factor varies also in the various ratings, ranging from .1300 to .3265. Sociability, neatness, intellectual originality, leadership, and self-assertion were rated with comparatively less chance error; while industry, carefulness, and independence were rated with more chance errors. As a whole, however, the chance errors are all quite small.

It is interesting to compare the loadings of the two general factors in each of the ratings. They are independent from each other. One rating may have a low loading of one of the two factors but a high loading of the other. For example, the rating on industry has a comparatively low loading of halo, while it has a comparatively high loading of will-power. The result is just what is to be expected. As industry is a trait which can be easily observed and specifically rated, it would be accordingly less affected by the halo. And, since industry is a result of will-power, naturally its rating is highly saturated with the general factor. These two factors, although both running through all of the ratings on the various traits, are distinct in nature. While the halo is a psychological error of a rater's judgment; the will-power belongs to the personality of the rated. The two general factors would come

out together, were it not for the statistical device of their separation, as discussed in Part II.

According to these data, there are three factors besides chance errors in each rating. One of the main purposes of this study was to show what proportion each factor plays in accounting for the total variance. On the average, as presented in Table XII, the halo factor accounts for 17.13 per cent, the general factor accounts for 31.59 per cent, the specific factors account for 46.53 per cent, and the chance errors account for 4.75 per cent. The halo is an inescapable error in rating method; but it appears not to be so large as it was commonly believed to be. This is probably due to the specific definition of the traits in this study. The chance errors are almost negligible. The general factor is probably as important as the specific factors for the understanding of a personality. The combination of both the general factor and the specific factor for the rating on a trait accounts for 78.12 per cent of the variance. Thus, the ratings are quite significant for the study of personality.

SUMMARY

According to these data, there are four factors operative in each rating, namely, the general factor of will-power, the specific factor, the halo, and the chance error.

The loading of each factor in each rating varies with the traits. On the average, the specific factor accounts for a little less than one-half of the total variance, the general factor of will-power accounts for about one-third, the halo accounts for about one-sixth, and the chance error accounts for about five per cent.

Since both the systematic error of the halo and the chance error are rather small and the specificity of the ratings is high, the ratings seem to be quite useful for the study of personality.

¹⁷ L. L. Thurstone, *Vectors of Mind*, University of Chicago Press, Chicago, 1936, p. 68.

